

CLAIMS

1. A system for automatically identifying a form or a page in multi-page form, comprising:

a digitizing pen;

a digitizing tablet, further comprising:

a support surface for supporting a form; and

means to detect pen stroke data when the digitizing pen is used to enter data in data entry fields on the form placed on the support surface, said pen stroke data having content information that is information requested by the form and location information that indicates the location on the form where the pen stroke data was entered; and

form selection means to select an electronic image of the form which data was entered on by the digitizing pen by selecting the best match of the pen stroke data with the electronic images of the forms;

whereby the system can be used to automatically identify the form being used based on the pen stroke data.

2. A system, as in claim 1, wherein:

the form selection means compares the location information from the pen stroke data with the location of data entry fields on forms, and selects the electronic image of the form related to the form on which the data was entered by selecting the best match of the location information with the location of the data entry fields on the electronic images of the forms;

whereby the electronic form image is selected by determining the location of data entered on the form.

3. A system, as in claim 2, wherein:

each form used by the system has data entry fields in different locations on the form; and

the form selection means distinguishes between forms by comparing the location information in the pen stroke data with the unique location of data entry fields on individual forms;

whereby the selection of an electronic image of a form is made more fault tolerant by placement of data entry fields on different forms in disparate locations.

4. A system, as in claim 2, wherein the form selection means further comprises:

means to calculate a data bounding box by identifying a discrete block of writing in the pen stroke data;

means to calculate a field bounding box for each data entry field on each form;

means to compare the distances between the corners of the data bounding box and the corners of the field bounding boxes; and

means to select the electronic image of the form which has the minimum distances between the corners of the data bounding box and the corners of the field bounding boxes;

whereby an electronic image of a form is selected based on the proximity and overlap of the data bounding box and the field bounding box.

5. A system, as in claim 4, further comprising:

means to select, when the data bounding box and the field bounding box do not overlap, the field bounding box which most closely matches the data bounding box;

whereby the system automatically compensates for form registration irregularities by selecting the best match of potential field bounding boxes for a given data bounding box.

6. A system, as in claim 4, further comprising:

means to limit initial comparisons of data bounding boxes and field bounding boxes to data bounding boxes and field bounding boxes in the same region of the form;

whereby the selection of an electronic image of a form is completed more rapidly by limiting the initial comparison of the data bounding box and the field bounding boxes to a preselected geographic area of the form.

7. A system, as in claim 4, further comprising:

means to limit comparisons of subsequent data bounding boxes, once a previous data bounding box and field bounding box have been matched, to field bounding boxes in the form if the subsequent data bounding box is located to the right or below the previous data bounding box;

whereby the selection of an electronic image of a form is completed more rapidly by temporarily limiting the comparison of subsequent data bounding boxes to field bounding boxes on the same form.

8. A system, as in claim 4, further comprising:

means to select the next page of a multi-page form as the initial candidate for comparison when a subsequent bounding box is above the previously matched data bounding box and field bounding box,;

whereby the selection of an electronic image of a form is completed more rapidly by comparing subsequent higher location data bounding boxes to field bounding boxes on the next page of a form.

9. A system, as in claim 4, further comprising:

means to select a particular electronic image of a form, when several forms are potential matches, by selecting the electronic image of the form with the most matches of data bounding boxes and field bounding boxes;

whereby the selection of an electronic image of a form is made by choosing the electronic image of the form with the most data bounding box/field bounding box matches.

10. A system, as in claim 4, further comprising:

means to determine the boundary of a data bounding box by setting the beginning of a data bounding box by detecting the initial pen stroke data, setting the end of a data bounding box by detecting the last pen stroke data, determining the vertical midpoint of the pen stroke data, and setting the height of the data bounding box to define an area above and below the midpoint of the pen stroke data which includes a substantial portion of the pen stroke data.

11. A system, as in claim 10, wherein:

the height of the data bounding box is approximately one standard deviation from the midpoint of the pen stroke data.

12. A system, as in claim 10, wherein:

the height of the data bounding box is approximately a fraction of, or a multiple of, a standard deviation from the midpoint of the pen stroke data.

13. A system, as in claim 4, further comprising:

means to generate a binary bitmap of data representing the pen stroke data;

comparison means to compare the binary bitmap of data from the pen stroke data with a binary bitmap of each page of an electronic image of a form, wherein the input fields of the electronic image of the form have no pixels and the non-input fields of the forms have pixels; and

selecting the best match between the pen stroke data and the electronic image of the form by identifying the form page that has the least number of overlapping pixels.

14. A system, as in claim 1, wherein:

the form selection means determines whether the content information from the pen stroke data contains content identifiable data unique to a particular form, and selects the electronic image of the form related to the form on which the data was entered when content identifiable data is present;

whereby the electronic form image is selected by detecting that content identifiable data was entered on the form.

15. A method of automatically identifying a form or a page in multi-page form, including the steps of:

using a digitizing pen and a digitizing tablet to generate pen stroke data when data is entered in data entry fields on a form; and

using the pen stroke data to select an electronic image of the form on which the pen stroke data was entered by selecting the best match of the pen stroke data with data entry fields on the electronic images of the forms;

whereby the system can be used to automatically identify the form being used based on the pen stroke data.

16. A method, as in claim 15, including the additional steps of:

including location information in the pen stroke data when the pen stroke data is generated;

comparing the location information with the location of data entry fields on electronic images of the forms to determine a best match; and

selecting the electronic image of the form based on the best match;

whereby the electronic form image is selected by determining the location of data entered on the form.

17. A method, as in claim 16, including the additional steps of:

placing the data entry fields on each form in different locations; and

selecting forms by comparing the location information in the pen stroke data with the unique location of data entry fields on individual forms;

whereby the selection of an electronic image of a form is made more fault tolerant by placement of data entry fields in disparate locations on different forms.

18. A method, as in claim 16, including the additional steps of:

identifying a discrete block of writing in the pen stroke data and calculating a data bounding box defining the discrete block of data;

calculating a field bounding box for each data entry field on each form; and

comparing the distances between the corners of the data bounding box and the corners of the field bounding boxes, and selecting the electronic image of the form which has the minimum distances between the corners of the data bounding box and the corners of the field bounding boxes;

whereby an electronic image of a form is selected based on the proximity and overlap of the data bounding box and the field bounding box.

19. A method, as in claim 18, including the additional steps of:

determining when the data bounding box and the field bounding box do not overlap; and

selecting the field bounding box which most closely matches the non-overlapping data bounding box;

whereby the system automatically compensates for form registration irregularities by selecting the best match of potential field bounding boxes for a given data bounding box.

20. A method, as in claim 18, including the additional step of:

limiting initial comparisons of data bounding boxes and field bounding boxes to data bounding boxes and field bounding boxes in the same region of the form;

whereby the selection of an electronic image of a form is completed more rapidly by limiting the initial comparison of the data bounding box and the field bounding boxes to a preselected geographic area of the form.

21. A method, as in claim 18, including the additional step of:

limiting comparisons of subsequent data bounding boxes, once a previous data bounding box and field bounding box have been matched, to field bounding boxes in the form if the subsequent data bounding box is located to the right or below the previous data bounding box;

whereby the selection of an electronic image of a form is completed more rapidly by temporarily limiting the comparison of subsequent data bounding boxes to field bounding boxes on the same form.

22. A method, as in claim 18, including the additional step of:

selecting the next page of a multi-page form when a subsequent bounding box is above the previously matched data bounding box and field bounding box;

whereby the selection of an electronic image of a form is completed more rapidly by comparing subsequent higher location data bounding boxes to field bounding boxes on the next page of a form.

23. A method, as in claim 18, including the additional step of:

selecting, when several forms are potential matches, the electronic image of the form with the most matches of data bounding boxes and field bounding boxes;

whereby the selection of an electronic image of a form is made by choosing the electronic image of the form with the most data bounding box/field bounding box matches.

24. A method, as in claim 18, including the additional steps of:

setting the left edge of a data bounding box at the approximate left edge of the pen stroke data;

setting the right edge of a data bounding box at the approximate right edge of the pen stroke data;

determining the vertical midpoint of the pen stroke data, and setting the height of the data bounding box to define an area above and below the midpoint of the pen stroke data which includes a substantial portion of the pen stroke data; and

setting the data bounding box to match the left edge, the right edge, and the height above and below the midpoint of the pen stroke data;

whereby the size and location of the data bounding box is determined.

25. A method, as in claim 24, including the additional step of:

setting the height of the data bounding box to approximately one standard deviation from the midpoint of the pen stroke data.

26. A method, as in claim 24, including the additional step of:

setting the height of the data bounding box to approximately a fraction of, or a multiple of, one standard deviation from the midpoint of the pen stroke data.

27. A method, as in claim 15, including the additional steps of:

detecting content information in the pen stroke data;

selecting the electronic image of the form related to the form on which the data was entered when the electronic image of the form is intended to hold corresponding content identifiable data;

whereby the electronic form image is selected by detecting that content identifiable data was entered on the form.

28. A method, as in claim 15, including the additional steps of:

generating a binary bitmap of data from the pen stroke data; and

comparing the binary bitmap of data from the pen stroke data with a binary bitmap of each page of an electronic image of a form, wherein the input fields of the electronic image of the form have no pixels and the non-input fields of the forms have pixels;

selecting the best match between the pen stroke data and the electronic image of the form by identifying the form page that has the least number of overlapping pixels.

29. A method of identifying a form from pen stroke data that is generated when filling out that paper form, including the steps of:

placing a paper form on a digitizing tablet;

capturing pen stroke data using a digitizing pen;

isolating the pen stroke data into groups separated by the time at which the marks were made, their location on the page of the form, and the proximity of the pen stroke data to other pen stroke data on the page;

matching each isolated group of pen stroke data to a field on one of a set of electronic image of the forms by minimizing the distance between the corners of a box delimiting the isolated group of pen stroke data and the corners of a box delimiting the fields on each electronic image of the form; and

selecting an electronic image of the form for which the combined distances between the corners of the group of pen stroke data and the fields are a minimum.

30. A method, as in claim 29, including the additional steps of:

optimizing the matching of each isolated group of pen stroke data by comparing the groups created at the earliest times with the first fields in the electronic image of the form;

comparing fields near the top of the following page with pen stroke data when the pen stroke data that immediately follow previously entered pen stroke data at the bottom of the form and overlap previously entered pen stroke data which have been matched to fields on the page of the form.

31. A method, as in claim 29, including the additional steps of:

examining isolated groups of pen stroke data to determine if the distance between them is less than a preselected numerical threshold; and

categorizing the pen stroke data as a single group of pen stroke data if the distance does not exceed the preselected numerical threshold, and categorizing the pen

stroke data as separate groups of pen stroke data if the distance exceeds the preselected numerical threshold.

32. A method, as in claim 29, including the additional steps of:

aligning the pen stroke data by defining a box delimiting the isolated group of pen stroke data;

rotating the pen stroke data used to generate the box to minimize the distance between its corners and the corners of a field delimiting box and applying this rotation to other groups of pen stroke data before computing their boundary boxes;

whereby the misalignment of the printed form upon the writing surface is corrected by rotating the pen stroke data prior to computing the boundary boxes to minimize distance between the corners of the boxes.

33. A method, as in claim 29, including the additional step of:

increasing the accuracy of form selection by ensuring that fields on different pages do not have exactly the same position and/or size on other pages.

34. A method, as in claim 29, including the additional steps of:

determining via handwriting recognition if pen stroke data contains content identifiable data; and

selecting an electronic image of a form based on the presence or absence of the content identifiable data.

35. A method, as in claim 29, including the additional steps of:

generating a binary bitmap of data from the pen stroke data;

comparing the binary bitmap of data from the pen stroke data with a binary bitmap of each page of an electronic image of a form, wherein the input fields of the electronic image of the form have no pixels and the non-input fields of the forms have pixels; and

selecting the best match between the pen stroke data and the electronic image of the form by identifying the form page that has the least number of overlapping pixels.

36. A method, as in claim 35, including the additional steps of:

using a boolean AND operation to compare the bitmaps of the pen stroke data and the electronic image of the form.